The Drosophila Molybdenum Cofactor Gene *cinnamon* Is Homologous to Three *Escherichia coli* Cofactor Proteins and to the Rat Protein Gephyrin

K. Puloma Kamdar, Michael E. Shelton and Victoria Finnerty

Department of Biology, Emory University, Atlanta, Georgia 30322

Manuscript received October 21, 1993
Accepted for publication March 10, 1994

ABSTRACT

Essentially all organisms depend upon molybdenum oxidoreductases which require a molybdopterin cofactor for catalytic activity. Mutations resulting in a lack of the cofactor show a pleiotropic loss of molybdoenzyme activities and thereby define genes involved in cofactor biosynthesis or utilization. In prokaryotes, two operons are directly associated with biosynthesis of the pterin moiety and its side chain while additional loci play a role in the acquisition of molybdenum and/or activation of the cofactor. Here we report the cloning of *cinnamon*, a Drosophila molybdenum cofactor gene encoding a protein with sequence similarity to three of the prokaryotic cofactor proteins. In addition, the Drosophila *cinnamon* protein is homologous to gephyrin, a protein isolated from the rat central nervous system. Our results suggest that some portions of the prokaryotic cofactor biosynthetic pathway composed of monofunctional proteins have evolved into a multifunctional protein in higher eukaryotes.

THE molybdoenzymes (except for nitrogenase) are characterized by an absolute requirement for a molybdopterin cofactor. These enzymes catalyze a variety of reactions in both prokaryotic and eukaryotic organisms. One such enzyme is assimilatory nitrate reductase (EC 1.6.6.1) which is involved in the uptake of soil nitrogen by higher plants. Global estimates suggest that more than 90% of all nitrogen assimilated by plants proceeds through this pathway (reviewed by WARNER and KLEINHOFS 1992). Sulfite oxidase (EC 1.8.2.1) is a mitochondrial intermembrane enzyme which catalyzes the terminal step in the catabolism of sulfur-containing amino acids and is also involved in the removal of sulfites (COHEN *et al.* 1973). Xanthine dehydrogenase/oxidase (EC 1.2.3.2) is found in most organisms and is involved in purine metabolism. In certain insects aldehyde oxidase (EC 1.2.3.1) is involved in regulating mating pheromone concentration (TASAYCO and PRESTWICH 1990) and in higher plants it catalyzes the last step in the biosynthesis of the stress hormone, abscisic acid (SINDHU *et al.* 1990). Virtually all organisms display some subset of molybdoenzyme activities and these enzymes, which play essential metabolic roles have been the object of intensive study.

The requirement for a molybdenum cofactor (MoCF) was originally predicted by genetic analysis. PATEMAN and co-workers identified five groups of Aspergillus mutations showing a pleiotropic loss of two molybdoenzyme activities, nitrate reductase and xanthine dehydrogenase. Since the mutants were found to contain enzymatically inactive apoenzymes, they hypothesized that these enzymes must require a hitherto unknown common molybdenum cofactor, and further predicted that their mutants identified loci involved in the biosynthesis or activation of the cofactor (PATEMAN *et al.* 1964). Such mutants define a group of loci referred to as molybdenum cofactor genes, which have been studied in several other organisms including Neurospora (NASON *et al.* 1971), *Escherichia coli* (AMY 1981; STEWART and MACGREGOR 1982), Arabidopsis (BRAAKSMA and FEENSTRA 1982), Chlamydomonas (AGUILAR *et al.* 1991), Drosophila (WARNER and FINNERTY 1981) and humans (JOHNSON *et al.* 1989). Although the prokaryotic loci have been cloned and specific functions assigned to several of the genes (reviewed by RAJAGOPALAN and JOHNSON 1992) there is no information concerning the structure of the eukaryotic molybdenum cofactor genes.

The eukaryotic MoCF is a novel 6-alkyl pterin in which the alkyl group is a four carbon side chain with a terminal phosphate and a 2,3-dithiolene coordinated to the molybdenum (RAJAGOPALAN and JOHNSON 1992). Biologically active MoCF is measured by an assay wherein fungal aponitrate reductase is provided with a possible source of MoCF which can associate with the apoenzyme to give nitrate reductase activity (NASON *et al.* 1971). Mutations showing the pleiotropic loss of molybdoenzyme activities may thus be classified into those lacking the cofactor, which are probably involved in the biosynthesis of the cofactor, and those which have cofactor, but may be involved in other processes such as molybdenum transport or activation. There are four Drosophila loci which show defects in molybdoenzyme activities: *cinnamon*, *low xanthine dehydrogenase*, *maroon-like* and *Aldox-2* (BAKER 1973; SCHOTT *et al.* 1986; FINNERTY 1976; MEIDINGER and BENTLEY 1986). Mutations for two of these loci, *cinnamon* and *low xanthine dehydrogenase*, show MoCF activity which is very low or absent (WARNER and FINNERTY 1981) and are therefore candidates for genes involved in cofactor biosynthesis.

This report describes the molecular cloning and characterization of the *cinnamon* region in *Drosophila melanogaster*. Evidence is presented which indicates that the region contains a gene that encodes a protein with striking similarities to three of the *E. coli* molybdenum cofactor proteins. In addition, the Drosophila cofactor gene product shows homology along its entire length to a protein isolated from the rat central nervous system. Our results suggest that a larger eukaryotic cofactor gene may have evolved to carry out several of the functions involved in cofactor synthesis. This apparent consolidation of functions which are carried out by separate proteins in *E. coli* has been observed in those few cases where both prokaryotic and eukaryotic homologs of biosynthetic genes are known.

MATERIALS AND METHODS

Drosophila strains and culture: We recovered the $cin^{\pi 1}$ mutant on an Oregon R background during a standard P-M dysgenic cross. The mutant is fertile (90% survival compared to wild type) for embryos derived from homozygous females (SHELTON 1990). The *Df(1)05-22-1/FM7,ct*^{5} /y^+ *Y* stock carries a spontaneous deficiency recovered by SCHALET (1986) from the Amherst wild-type strain. The *cin* alleles not specifically discussed here are described in STIVALETTA *et al.* (1988) and LINDSLEY and ZIMM (1992).

Flies were maintained at 25° on standard cornmeal media (ASHBURNER 1989) supplemented with mold inhibitor (Carolina Biological). The *cinnamon* mutants in our collection have very low but not totally absent levels of xanthine dehydrogenase (XDH) activity (STIVALETTA *et al.* 1988). As with other mutations affecting XDH, very low activities or even maternal effects lead to a normal eye color instead of the dull reddish brown characteristic of XDH-negative flies. To suppress low levels of XDH and permit scoring the *cin* mutant eye color, flies were reared on media containing an XDH inhibitor, allopurinol (0.015 μg/ml, final concentration) (Sigma A8003) (KELLER and GLASSMAN 1965).

Isolation of *cinnamon* genomic DNA and cDNAs from *D. melanogaster*: A Charon 4A library of Canton S DNA (MANIATIS *et al.* 1978) was screened to obtain genomic clones of *cinnamon*. Approximately 6×10^4 recombinant phage were screened by the method of BENTON and DAVIS (1977), using the TG-11 insert as a probe. The λgt10 2–12-hr embryonic cDNA library, provided by L. KAUVAR, was screened with genomic DNA inserts from the *cin* region. Lambda DNA was isolated from purified recombinant phage in large cultures by the CsCl gradient method of YAMAMOTO *et al.* (1970).

Plasmid subclones were generated from the recombinant phage clones by ligating gel-purified restriction fragments into the vector, pBluescript KS+. (Stratagene). Plasmid DNA was isolated by the alkaline lysis procedure of BIRNBOIM and DOLY (1979).

DNA sequencing and analysis: DNA sequencing was performed by the SANGER *et al.* (1977) dideoxy method using ^{35}S-labeled dATP and the Sequenase 2.0 Kit (U.S. Biochemical Corp.). For Bluescript clones, the sequencing was performed in conjunction with the M13 Universal or Reverse primers or custom synthesized 17- and 18-base oligonucleotide primers complementary to the sequence. The sequence was confirmed by sequencing both strands. Some genomic sequencing was performed using the CircumVent Thermal Cycle Dideoxy DNA sequencing kit (New England Biolabs). DNA sequence was compiled using GenePro (Riverside Scientific) and analyzed using the University of Wisconsin Genetics Computer Group (UWGCG) package (Version 7). The *cinnamon* DNA sequence has been submitted to GenBank (accession no. L19876).

Construction and screening of $cin^{\pi 1}$ genomic library: A $cin^{\pi 1}$ EMBL3 phage library was constructed with Stratagene's Lambda EMBL3/*Bam*HI vector kit according to MANIATIS *et al.* (1982). Recombinant phage were packaged using Stratagene Gigapack Gold *in vitro* packaging system. The unamplified $cin^{\pi 1}$ library was screened with the 3.8-kb *Eco*RI fragment of MES-1, according to MANIATIS *et al.* (1982).

Southern analysis: Genomic DNA was extracted from adult flies according to LIVAK's (1984) method. DNA was digested with restriction enzymes, separated by electrophoresis in 0.7% agarose gels and transferred onto Zeta-Probe membrane using an alkaline blotting procedure (BRL). DNA probes were labeled using the random priming method of FEINBERG and VOGELSTEIN (1983) with [^{32}P]dCTP (Amersham). Southern blots were hybridized according to the Zeta-Probe protocol at 42° and washed for 15 min at room temperature in 2 × SSC, 0.1% sodium dodecyl sulfate (SDS) followed by 0.1 × SSC, 0.1% SDS at 55° for 30 min. Autoradiography was performed at −70° using X-Omat AR X-ray film (Kodak).

Northern analysis: Flies were harvested at the indicated developmental stages and frozen immediately in liquid nitrogen. For embryo collections, adult flies in population cages prelayed for 1 hr and then were allowed to lay for 3 hr on Petri plates containing an apple juice/agar medium covered with yeast paste (ASHBURNER 1989). The plates were removed and incubated for the appropriate time interval at 25°. Pupae were hand collected from the walls of the culture container. Total RNA was purified from fly tissues according to the protocol of ARTAVANIS-TSAKONAS *et al.* (1983) and resuspended in diethyl pyrocarbonate-treated water. Poly(A)$^+$ RNA was bound to oligo-dT cellulose (equilibrated in binding buffer), washed with 0.15 M LiCl, 10 mM Tris, 1 mM EDTA, 0.1% SDS, and eluted with 2 mM EDTA, 0.1% SDS. RNA was precipitated with ethanol and stored at −70°. Concentration was determined by A_{260} readings.

RNA was subjected to electrophoresis on 1.2% agarose-formaldehyde gels in 10 × 4-morpholinepropanesulfonic acid buffer. RNA size standards (BRL) were used for molecular weight determination. Following electrophoresis, the gel was soaked in 20 × SSC for 45 min at room temperature, and the RNA was transferred onto Zeta-Probe membrane (BRL) in 10 × SSC. Probes were *in vitro* transcribed RNA from a Bluescript subclone following Stratagene instructions. Hybridization was carried out at 42°, and filters were washed at room temperature in 1 × SSC, 0.1% SDS for 15 min (2 washes), followed by 0.25 × SSC, 0.1% SDS at room temperature (2 washes). Autoradiography was performed as above.

RESULTS

Cloning of the *cinnamon* genomic region: The first *cinnamon* (*cin*) mutation was selected as a female sterile and placed distal to *yellow* (*y*:1,0.0) (BAKER 1973). Further cytogenetic analysis established that the (telomere to centromere) order of genes in this region is *lEC1* (also called *dmd* or *pch*), *lEC2*, *lJ1*, *cin*, *ewg*, *arth*, *y* (LEFEVRE 1981; FLEMING *et al.* 1989; LINDSLEY and ZIMM 1992). Using TG-11, a clone derived from a molecular analysis of *erect wing* (*ewg*), generously provided by R. J. FLEMING and K. WHITE, six phage were selected

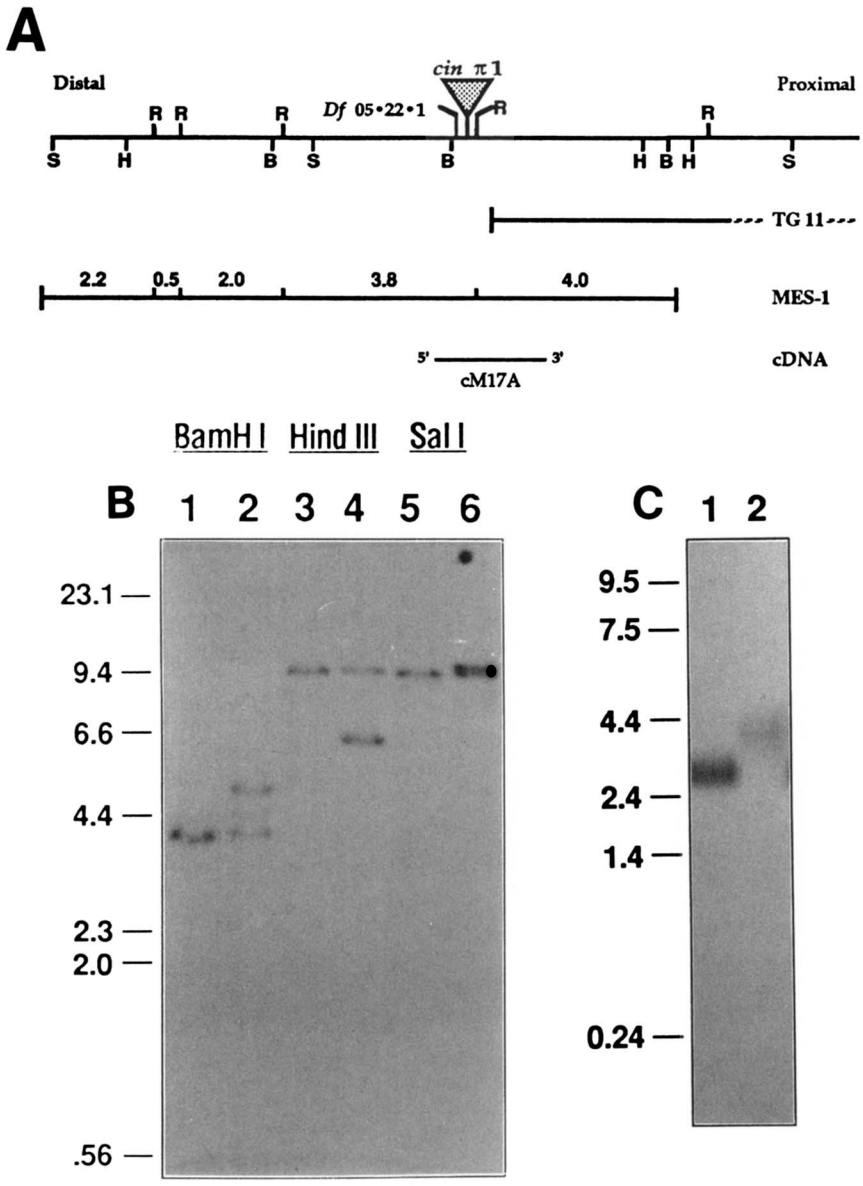


FIGURE 1.—Identification of sequences related to *cinnamon* function. (A) Organization of the *cinnamon* region. The horizontal line depicts 20 kb of genomic DNA for the *cin* region, the rightmost *Eco*RI site being approximately 92 kb distal to *yellow* on BEISSMAN's (1985) scale. Restriction enzyme sites are: H, *Hin*dIII; B, *Bam*HI; S, *Sal*I; R, *Eco*RI. MES-1 was obtained from the MANIATIS library using the probe TG-11, from the proximal adjacent *erect wing* region. The MES-1 insert was used to select cM17A from an embryonic cDNA library. The precise location depicted for cM17A and the $cin^{\pi 1}$ insertion was determined by sequence analysis. (B) Genomic Southern analysis of the proximal breakpoint of *Df(1)05-22-1*. Lanes 1, 3 and 5 contain genomic DNA from Amherst wild-type flies restricted as indicated. Lanes 2, 4 and 6 are *Df(1)05-22-1* heterozygous with Amherst, the parental strain of *Df(1)05-22-1*. The probe, a 650-bp fragment from the *Eco*RI site to the 5′ end of cM17A, reveals unique fragments with both *Bam*HI and *Hin*dIII digests indicating that the proximal breakpoint of *Df(1)05-22-1* falls within the region covered by the probe. (C) Northern analysis of the insertion mutant $cin^{\pi 1}$ results in the production of an mRNA approximately 400 bp larger then the 2.6-kb wild-type mRNA. Aliquots (10 μg) of poly(A)$^+$ wild-type (lane 1) and $cin^{\pi 1}$ (lane 2) RNA were probed with [^{32}P]rUTP-labeled RNA made from cM17A.

from the MANIATIS library. One phage, MES-1, included DNA distal to that of TG-11 and is depicted in Figure 1. MES-1 corresponds approximately to 156–169 on the CAMPUZANO *et al.* (1985) scale, roughly 92 kb distal to the *yellow* locus.

Evidence that MES-1 contains sequences corresponding to the *cin* locus was derived from Southern analysis of several *cin* mutants. One allele, $cin^{\pi 1}$, which we recovered via hybrid dysgenesis, revealed a 400-bp insertion when restricted with *Bam*HI and *Eco*RI and probed with the 3.8-kb *Eco*RI fragment of MES-1 (data not shown). Additional data suggesting that MES-1 represents DNA involved in *cinnamon* function came from the terminal deficiency, *Df(1)05-22-1*, which was known to complement *ewg*, but failed to complement *l(1)EC2*. This deficiency also failed to complement *cin* since heterozygous females of the type *Df(1)05-22-1/* cin^{MD} or *Df(1)05-22-1/*cin^{C23} display a brownish eye color characteristic of *cin* mutants that lack xanthine dehydrogenase activity (SHELTON 1990). Southern analysis (Figure 1B) reveals that the proximal breakpoint of *Df(1)05-22-1* falls within the 3.8-kb *Eco*RI fragment of MES-1.

One cDNA is directly associated with *cinnamon*: To focus on a possible *cin*-coding region, the 12.5-kb MES-1 insert was used to select a group of cDNAs. The position of the $cin^{\pi 1}$ insertion as well as the proximal breakpoint of *Df(1)05-22-1* served to focus attention on one of the cDNAs, cM17A (depicted in Figure 1A) which hybridized to both the 3.8- and 4.0-kb *Eco*RI fragments of MES-1. To determine whether cM17A might represent at least part of the *cin*-coding sequence, it was used to probe northern blots containing mRNA from various *cin* mutants. Figure 1C shows that the insertion mutant, $cin^{\pi 1}$ produces an mRNA which is approximately 400 bp larger than the 2.6-kb mRNA originating from the

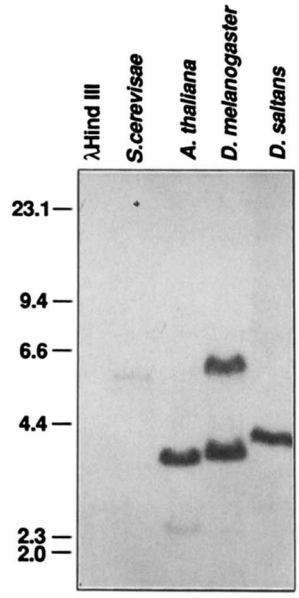


FIGURE 2.—Southern analysis suggests conservation of the cM17A sequence. Aliquots of 1.5 μg of *Saccharomyces cerevisae*, 3 μg of *A. thaliana*, 5 μg of *D. melanogaster* and *D. saltans* DNA were probed with the 1.5-kb *Eco*RI fragment from the carboxy-terminal end of cM17A. *S. cerevisae* DNA was restricted with *Bam*HI and all other DNA was digested using *Eco*RI. The two bands of hybridization in the *D. melanogaster* lane reflects a polymorphism for the most proximal *Eco*RI site depicted in Figure 1A.

parental chromosome. Although insertions often lead to the production of a truncated message, instances of larger messages have been reported (MOERMAN *et al.* 1988). The insertion in $cin^{\pi 1}$ is rather small and may lack a fortuitous transcription termination signal.

Southern analysis suggests conservation of the cM17A sequence among eukaryotes: Since the structure of MoCF appears to be identical in all eukaryotic organisms (KRAMER *et al.* 1987) there is the possibility that certain protein domains which bind the MoCF or its precursors might be conserved even at the nucleotide level. To examine this possibility DNA blots from various species were prepared to contain an approximately equal number of genome equivalents. An example of such a blot is shown in Figure 2. The blots were probed with cM17A and treated under conditions which would require conservation of at least a stretch of 100 bp of DNA in order to detect hybridization. There is clearly a healthy band of hybridization in *Drosophila saltans*, which has evolved independently of *D. melanogaster* for a sufficient period to allow nearly complete divergence of unconstrained sequences (50 million years; BEVERLEY and WILSON 1984). Also, there are bands detectable in *Arabidopsis thaliana* as well as a weak band in yeast. Under these conditions there is no hybridization in *E. coli* or in Streptomyces (data not shown). We take this as one piece of evidence to suggest that portions of the *cinnamon* gene may be conserved in eukaryotes.

Nucleotide sequence of *cinnamon*: The nucleotide and deduced amino acid sequences of the *cinnamon* coding region, including introns and 5′- and 3′-flanking regions is presented in Figure 3. A proposed heptanucleotide ATCAGTG transcription initiation site, present at position −236 of cM17A, is similar in 6 of the 7 bases to the reported insect consensus ATCA(G/T)T(C/T) for non-heat shock genes (HULTMARK *et al.* 1986). A possible TATA box is located at position −297 and the common polyadenylation signal AATAAA is found in the 3′ region of the cDNA. As with many eukaryotic genes, the longest open reading frame follows the first ATG of *cinnamon*, which is flanked by a purine at −3 and a G residue at +4 (KOZAK 1984). If this site is used for the initiation of translation, the transcript will have a 236-bp untranslated leader sequence and will generate a 601-amino acid polypeptide of relative molecular mass 66 kD. The sequence contains six introns all of which have consensus splice acceptor and splice donor nucleotide sequences (JACKSON 1991).

The 2.6-kb *cin* message is developmentally regulated: Several of the *cin* alleles are female sterile (PADILLA and NASH 1977; SHELTON 1990). Since the sterility can be rescued by a paternally derived cin^{+} allele, *cin* is classified as a maternal effect zygotic lethal (PERRIMON *et al.* 1989). Thus it was of interest to examine the developmental appearance of the 2.6-kb mRNA (Figure 4). The 2.6-kb transcript is expressed predominantly during the embryonic stages, appears to diminish during the larval stages and finally becomes abundant again during adulthood. There is a suggestion of a slightly smaller maternal message, but as yet this possibility has not been rigorously examined.

Amino acid sequence similarities to three *E. coli* MoCF proteins: To examine the function of the *cin* product, we searched for similarities of the deduced amino acid sequence to characterized gene products. Database comparisons (Figure 5) of the conceptual translation of the cM17A coding sequence reveal significant sequence similarities with the *E. coli* molybdenum cofactor proteins chlE, chlA2 and chlG. As illustrated in Figure 6, the first 190 residues at the amino terminus are homologous to the *E. coli* MoCF proteins chlA2 and chlG. It had not been previously recognized that these two *E. coli* proteins are also homologous to one another, showing 52% similarity. The chlG protein appears to play some role in the intracellular utilization of Mo since labeling studies associate it with a high molecular weight molybdenum binding protein (S. M. HINTON, personal communication). The chlA2 protein has a motif found in folate-containing molecules, some of which are pterin-binding proteins (RIVERS *et al.* 1993). The *chlA2* gene is not currently represented by mutations and resides on the *chlA* operon, which contains other open reading frames associated with MoCF biosynthesis (Figure 7). The high degree of similarity

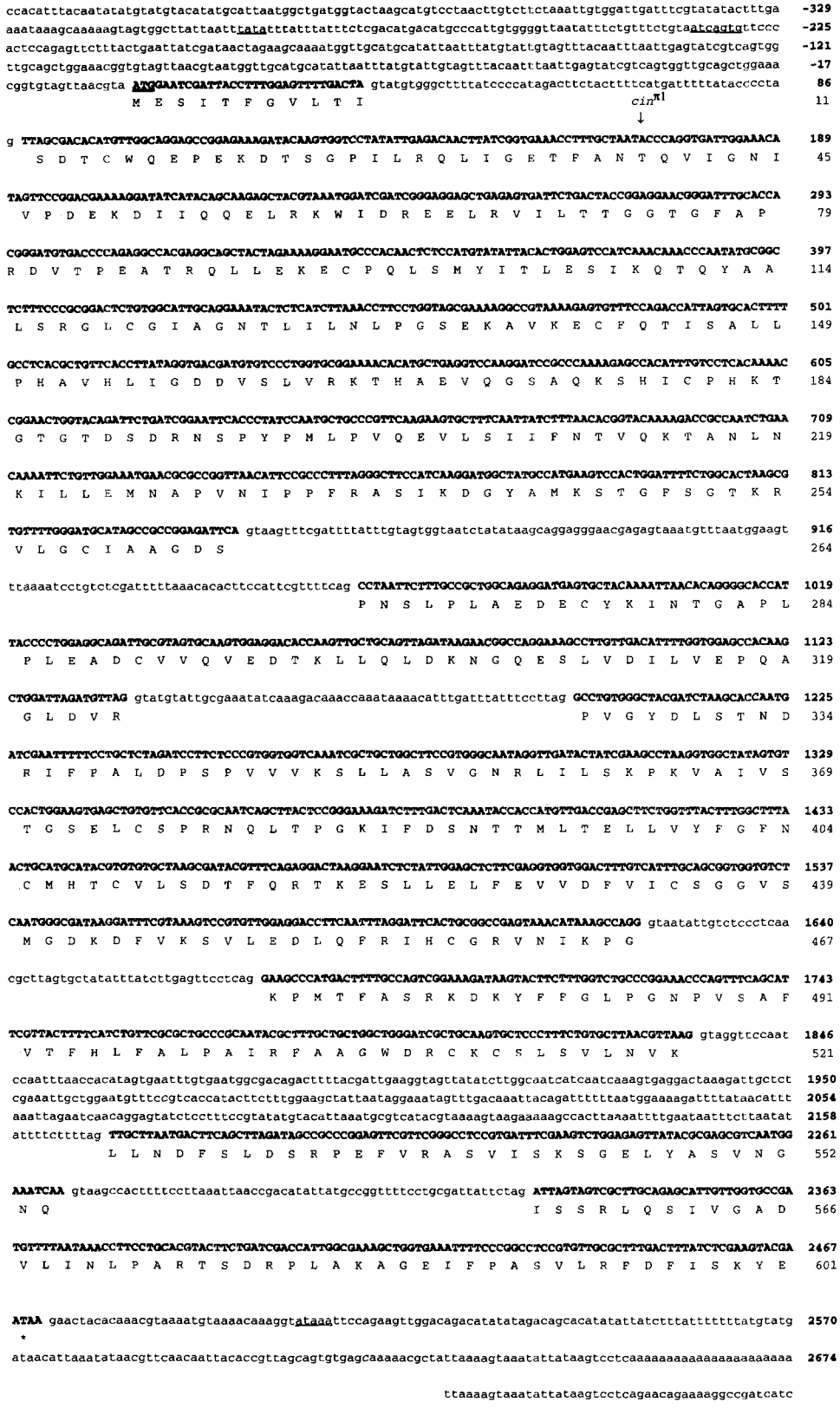


FIGURE 3.—Nucleotide and deduced amino acid sequence of *cinnamon*. The complete sequence of the coding region (uppercase) and intronic as well as 5′- and 3′-flanking genomic sequences (lowercase) is presented. The amino acid sequence (in one letter code) is also shown. Both strands of cM17A were sequenced and corroborated by sequencing the genomic region. The upstream consensus transcriptional start site and a possible TATA box are underlined. Primer extension analysis (data not shown) suggests that the transcriptional start site at −297 is utilized. The potential translational start site is double underlined. A putative polyadenylation signal in the 3′ region is indicated by underlining. The 400-bp insertion (an internally deleted *P* element) in the mutant allele cin^{m1} is indicated above the sequence. The full length Drosophila *cinnamon* sequence has been deposited with GenBank (accession no. L19876).

between chlG and chlA2 extends essentially over their entire length suggesting that they may carry out similar functions or they may associate with the same or similar molecules. Interestingly, attempts to mutagenize *chlA2* have not yet been successful, possibly because the chlA2 function and the chlG function could be at least partially redundant and chlG may substitute well enough to allow *chlA2* mutants to escape selection. The similarity of the

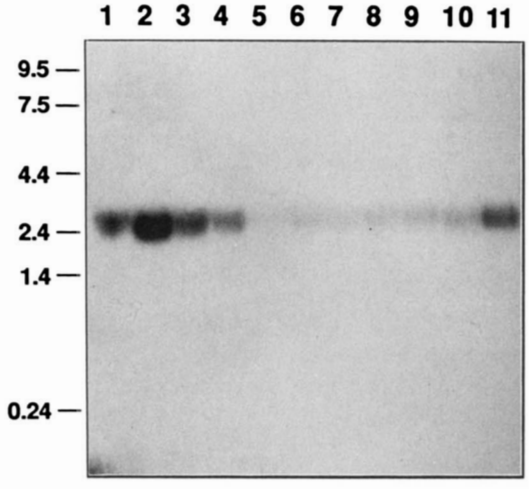


FIGURE 4.—Developmental expression of the 2.6-kb *cin* mRNA. Each lane contains 10 μg of poly(A)$^+$ RNA from the following stages: (1) 0–3-hr embryos; (2) 3–8-hr embryos; (3) 9–15-hr embryos; (4) 15–24-hr embryos; (5) first instar larvae; (6) second instar larvae; (7) third instar larvae; (8) 0–24-hr pupae; (9) 0–48-hr pupae; (10) 48–96-hr pupae; (11) adult. The probe was a single stranded RNA made from cM17A.

cinnamon protein to chlA2 (58%) and to chlG (54%) suggests the possibility that the amino portion of cinnamon (which appears to have consolidated chlA2 and chlG) could act as a single domain which channels or presents MoCF precursors to the carboxy domain or even to other proteins in the pathway. Another possibility would be that the amino portion of cinnamon carries out two separate functions associated with MoCF biosynthesis.

Non-overlapping with this amino terminal homology, the carboxy two-thirds of cinnamon is 26% identical and 47% similar to the prokaryotic chlE (Figure 6). The *chlE* gene (NOHNO *et al.* 1988) can be tentatively placed in the MoCF biosynthetic scheme as depicted in Figure 7. Studies of MoCF biosynthesis have identified two precursors, both containing pterin and the phosphorylated side chain. One, called "molybdopterin" has the dithiolene but lacks Mo; the other, precursor Z, lacks both Mo and the dithiolene moiety. Three genes (*chlA4*, *chlA5* and *chlN*) were first implicated in the conversion of precursor Z because mutants accumulate precursor Z and lack molybdopterin and the MoCF (JOHNSON and RAJAGOPALAN 1987). The most closely studied *chlE* mutant is *chlE5*, which has precursor Z but only 6% of wild-type levels of molybdopterin and no MoCF activity (JOHNSON and RAJAGOPALAN 1987). This suggests to us that the chlE function is involved in the converting step and that *chlE5* may be a partially active allele. Yet recent studies show that all that is necessary for the *in vitro* conversion of precursor Z to molybdopterin are the three proteins encoded by *chlA4*, *chlA5* and *chlN* (PITTERLE *et al.* 1993). One possibility is that while the chlE function is not required *in vitro* it may be needed *in vivo* to stabilize the converting factor multimer. The fact that the extreme lability of both precursor Z and molybdopterin to oxidative damage is ameliorated *in vitro* when they are bound to the converting factor (PITTERLE *et al.* 1993) argues for the necessity of proteins to stabilize pterin-based precursor molecules *in vivo*. The observation that areas of amino acid identity between chlE and cinnamon essentially cover the entire length of chlE suggests that the two are not simply sharing a structural motif, but rather that this domain of the cinnamon protein may have a similar function in eukaryotes.

Cinnamon is homologous to the rat protein gephyrin: The cinnamon protein is 39% identical and 62% similar to gephyrin over its entire length and both proteins appear to be composed of two domains (Figures 5 and 6). Gephyrin copurifies with the glycinergic receptor and also binds polymerized tubulin (PRIOR *et al.* 1992). Gephyrin is located in brain tissue on the cytosolic face of postsynaptic membranes and functional evidence indicates that it plays a role in linking the glycinergic receptor to the subsynaptic cytoskeleton (KIRSCH *et al.* 1993a). Five alternatively spliced gephyrin mRNAs are found in various areas of the brain suggesting that it may interact with several different proteins (KIRSCH *et al.* 1993b). Gephyrin has 57% similarity at the carboxy terminus to *chlE* which has been noted before (PRIOR *et al.* 1992). Interestingly, gephyrin also shares sequence similarity at its amino terminus with chlA2 (59%) and chlG (65%). As tabulated in Figure 6 both eukaryotic proteins show approximately the same degree of similarity to the prokaryotic MoCF genes. Moreover the alignment in Figure 5 clearly illustrates the great similarity between gephyrin and cinnamon and the prokaryotic MoCF proteins and shows homology extending essentially over the entirety of the *E. coli* proteins, rather than being confined to a few highly conserved motifs (Figure 5). This suggests that gephyrin is the mammalian homolog of *cinnamon*.

DISCUSSION

We have presented evidence for the molecular isolation of the cinnamon locus based on genomic mapping of the $cin^{\pi 1}$ allele and the proximal breakpoint of *Df(1)05-22-1* to the locus. In addition, we also detect an aberrant sized transcript in the insertional $cin^{\pi 1}$ allele. Database comparisons of the deduced amino acid sequence of cinnamon reveals extensive homology with a rat protein which links glycine receptors to microtubules. In addition, three prokaryotic MoCF genes, chlA5, chlG and chlE show a striking similarity to cinnamon. Cinnamon shares common residues with chlG and chlA5 at its amino terminal end. Although they show significant sequence similarity, chlA5 and chlG appear to have different roles in the synthesis of the MoCF. In light of the homology found between these two proteins, it seems likely that they associate with similar molecules,

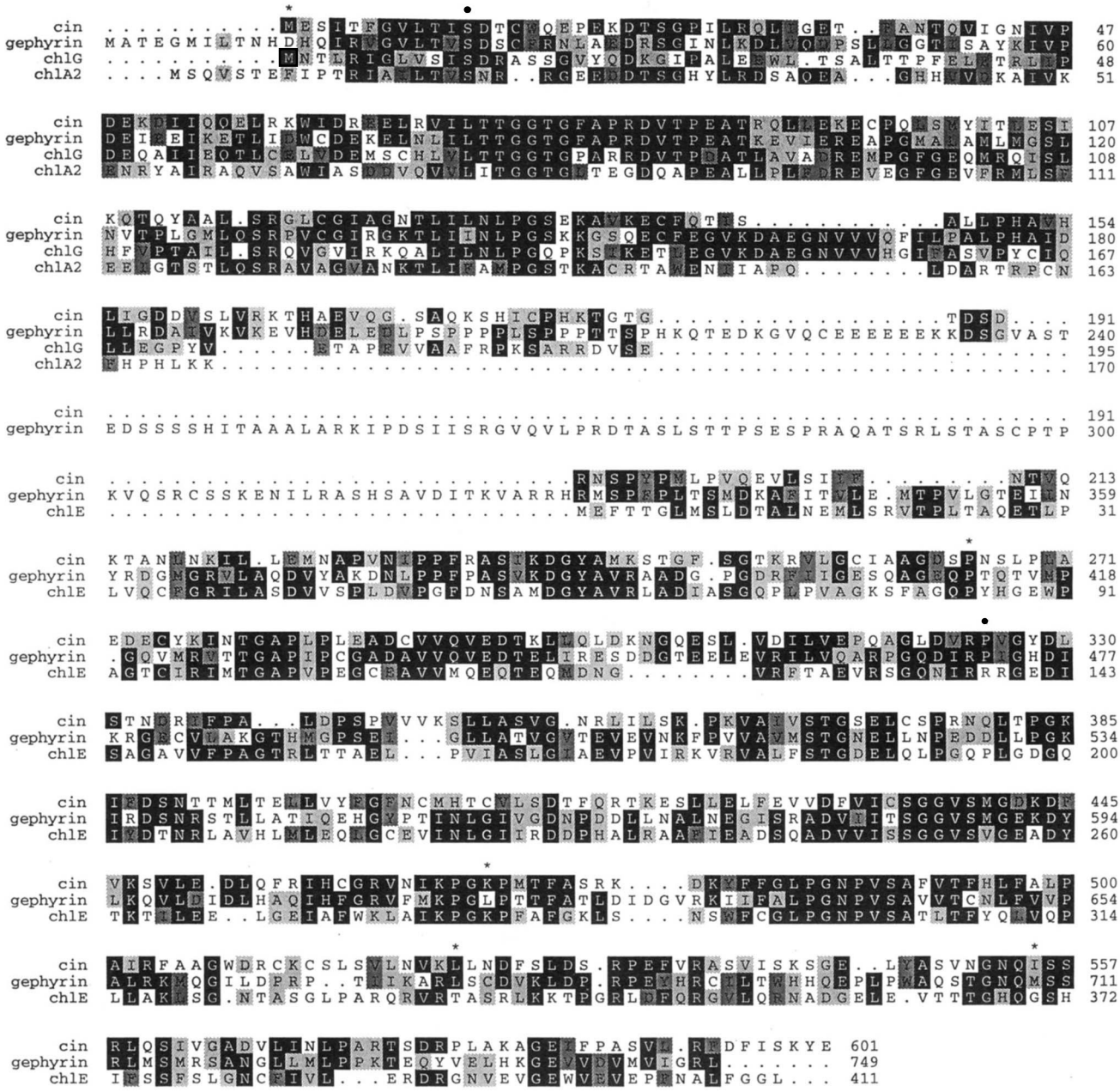
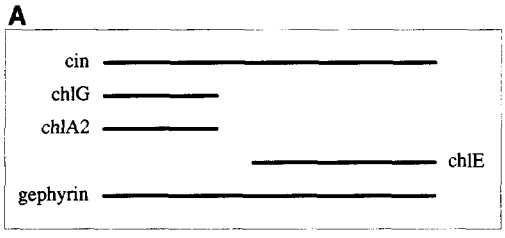


FIGURE 5.—Amino acid sequence similarity between *cinnamon*, the *E. coli* molybdenum cofactor genes and the rat protein gephyrin. Black boxes show sequence identity whereas dark gray and light gray shading represents more and less favorable conservative substitutions respectively. Gaps introduced in a sequence to optimize the alignment are represented by dots. Sequence similarities were initially determined from searches conducted at the NCBI using the BLAST network service. Full alignments using a gap weight of 3.0 were made with the multiple sequence analysis program from the Genetics Computer Group Inc. (Version 7, 1991) Madison, Wisconsin, and the Prettybox program by RICK WESTERMAN. The symbol comparison table for proteins was based on the Dayhoff PAM-250 matrix (SCHWARTZ and DAYHOFF 1979). The thresholds for the pairwise matching were 1.5 for identical, 1.0 for similar (dark gray boxes) and 0.5 for somewhat matching (light gray boxes). The seven exons of the cinnamon protein are indicated by asterisks at the beginning of each exon.

perhaps the MoCF precursor, at different stages in the pathway. At the carboxy-terminal end, cinnamon shows sequence similarity to chlE, also functioning in MoCF synthesis, perhaps to stabilize the converting factor. Thus it appears that *cin* has evolved to encode a multifunctional protein. The clustering of related biosynthetic steps into a multifunctional eukaryotic protein has been noted in other ancient essential pathways including purine biosynthesis as well as those for pyrimidines and fatty acids (HENIKOFF *et al.* 1986; DAVIDSON *et al.* 1993; AMY *et al.* 1992). DAVIDSON *et al.* (1993) put forth the argument that there are at least two mechanisms acting as the driving force for the creation of eukaryotic multienzymatic proteins: channeling and the selective advantage of coordinate gene and protein expression. The prokaryotic MoCF genes are encoded at two oper-



B

	cin	gephyrin	chlG
gephyrin	39 / 62		
chlG	27 / 54	37 / 65	
chlA2	27 / 58	27 / 59	22 / 52
chlE	26 / 47	35 / 57	

FIGURE 6.—Regions of homology and summary of the pairwise comparisons between cinnamon, gephyrin, chlA2, chlG and chlE. (A) Relationship of the homologous proteins. (B) The percent amino acid identity followed by their similarity (identities plus conservative substitutions) is shown. In all pairwise comparisons the smaller protein was used as the denominator.

ons, thus allowing the coordinate expression of these genes. This would significantly reduce the selective pressure for the fusion of genes encoding monofunctional proteins, allowing bacteria to maintain separate genes for these enzymes. However, in eukaryotes, the pressure to coordinately express MoCF genes might have led to the evolution of a protein with multiple enzymatic domains. It is also plausible that the necessity to channel the product of one step in the MoCF pathway to the next step acted as a strong selective advantage leading to a multienzymatic protein. Perhaps the amino terminal portion of *cin* acts to bind a MoCF precursor presenting it to other proteins involved in the synthesis of a complete MoCF.

Another reason for suspecting that the *cin* gene encodes a multifunctional protein is the complex complementation pattern seen at this locus (STIVALETTA *et al.* 1988). Various *cin* mutant heterozygotes show wild type and in some cases more than wild-type levels of complementation. The $cin^{\pi 1}$ allele, which has an insertion of 400 bp of *P* element DNA near to the 5′ end, is capable of producing a truncated protein (Figure 3). Interestingly, $cin^{\pi 1}$ shows very low levels of complementation with other *cin* mutants (our unpublished observations). One allele, cin^{C44}, which is now known to be a null for the 2.6-kb message, is completely noncomplementing. It will be very informative to examine the molecular localization of other *cin* alleles to determine if they affect different domains in the cinnamon protein and are therefore able to complement other mutants to wild-type levels.

If a possible model for the evolution of the *cin* gene involved intron-intron mediated rearrangements, then one might expect to find introns within the DNA sequences encoding each of the bridges separating the functional domains as one does for the CAD gene, which is involved in pyrimidine biosynthesis in higher eukaryotes (DAVIDSON *et al.* 1993). On the other hand, organisms such as Drosophila whose genomes have been streamlined through selection, may show an absence of such introns. A puzzling feature of the *cinnamon* gene is its six introns (Figure 4), more than expected for a Drosophila protein of this size. Surprisingly, there is no intron separating the two domains in *cin*. As yet no correlation can be made between the exon boundaries of *cin* and the homology boundaries of the *E. coli* proteins. The position of gephyrin introns is unknown.

The *cinnamon* gene is clearly a MoCF gene, but its specific role (as well as that of its prokaryotic homologs has yet to be determined. Although PRIOR *et al.* (1992) noted gephyrin's homology to chlE, they have not commented on the chlG and chlA5 homologies and their interpretation of gephyrin function is strictly focused on a role in the nervous system. The homology of cinnamon and all three prokaryotic MoCF proteins to gephyrin extends essentially over the entire gephyrin protein suggesting that gephyrin is a mammalian homolog of *cinnamon*. The amino terminal portion of gephyrin contains several splice variants suggesting possible associations with different proteins. The numerous splice variants plus the fact that gephyrin is also found in several non-neural rat tissues such as kidney, liver and lung indicates that gephyrin plays a much more general role in microtubule-associated interactions (PRIOR *et al.* 1992). We would certainly expect that the MoCF pathway would operate in organs such as the liver where xanthine oxidase, sulfite oxidase, and aldehyde oxidase activities are also found. The carboxy-terminal region of gephyrin is constant suggesting that this region of the protein binds microtubules (FROEHNER 1993). Gephyrin also has a carboxy-terminal microbody motif suggesting that it could gain entry into membrane-enclosed compartments, such as a peroxisomes. Such a localization would not be surprising since xanthine dehydrogenase/oxidase is also a peroxisomal activity. One explanation for their apparently paradoxical functions is that one or both species could contain a few similar genes. Yet at present there is no reason to believe that either is a member of a structurally similar gene family which has functionally divergent members.

A second possibility is that the cinnamon/gephyrin protein actually carries out two entirely separate functions. Interestingly, many *cin* alleles are female sterile because their eggs do not support zygotic development. The probable proximal cause of the lethality

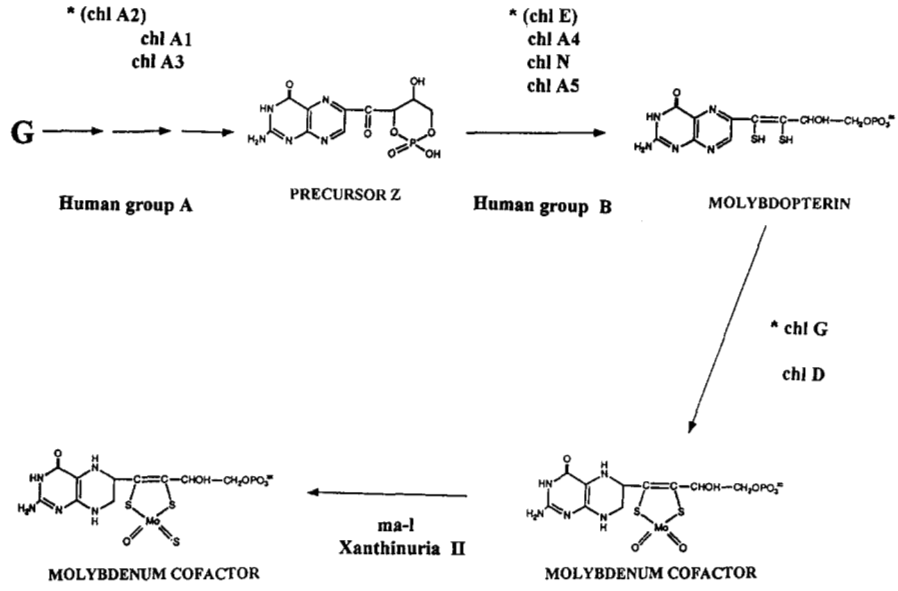


FIGURE 7.—Molybdenum cofactor biosynthesis. Labeling studies indicate that the pterin ring and side chain are derived from guanosine (WUEBBENS and RAJAGOPALAN 1993a) by a currently unknown series of reactions, possibly involving *chlA1*, *chlA2* and *chlA3*. Precursor Z has been isolated (WUEBBENS and RAJAGOPALAN 1993b) and can be converted *in vitro* to molybdopterin by the purified products of *chlA4* and *chlA5* which form a heteromultimeric converting factor (RAJAGOPALAN and JOHNSON 1992). The *chlN* product provides sulfur to the *chlA4* subunit (PITTERLE *et al.* 1993). The details of how the converting factor-molybdopterin complex acquires Mo *in vivo* are unknown. The *chlG* gene appears to be necessary for the efficient incorporation of Mo into Mo-enzymes and Mo-binding proteins (HINTON and DEAN 1990). The *chlD* operon is involved in molybdenum uptake by the cell. The function of the *chlA2* gene is unknown. Another gene, *chlB* (not depicted here) acts later in the pathway to mediate the acquisition of a phosphate-linked nucleotide found only in the prokaryotic forms of the MoCF (JOHNSON *et al.* 1990). The *E. coli* loci *chlE* and *chlN* have been renamed *moeA* and *moeB*; the *chlA1*, *A2*, *A3*, *A4* (or *M*) and *A5* are now moaA, B, C, D and E (SHANMUGAM *et al.* 1992). Human patients carrying the autosomal recessive molybdenum cofactor defect fall into two complementation groups based on the presence of precursor Z and the ability to carry out the *in vitro* conversion of precursor Z to molybdopterin (JOHNSON *et al.* 1989). The Drosophila *ma-l* gene is probably involved in the acquisition of the terminal sulfur ligand found in certain molybdoenzymes (WAHL *et al.* 1982). Human patients suffering from the inherited defect xanthinuria II, also appear to lack the function which carries out sulfuration of the terminal sulfur ligand (SIMMONDS *et al.* 1994). The three loci bearing homology to *cinnamon* are indicated with asterisks. Parentheses indicate loci whose position in the pathway is unclear at present.

observed in *cin* mutants is their sulfite oxidase defect (STIVALETTA *et al.* 1988). The biochemical phenotype of *cin* mutants closely resembles that of human patients suffering from the molybdenum cofactor defect, which is also proximally associated with a deficiency of sulfite oxidase activity. However, the primary clinical finding in human patients is a lack of development of the brain and nervous system (JOHNSON and WADMAN 1989). Examining the relationship of cinnamon and gephyrin and determining if these proteins actually carry out two functions may offer a clue to the exact cause of these developmental abnormalities.

This work was supported by a grant from the National Science Foundation, DCB 89169. K.P.K. received support from Sigma Xi. We thank D. H. BOXER and S. M. HINTON for sharing unpublished data and to our colleagues for helpful comments on the manuscript.

LITERATURE CITED

AGUILAR, M. R., J. CARDENAS and E. FERNANDEZ, 1991 Regulation of molybdenum cofactor species in the green alga *Chlamydomonas reinhardti*. Biochim. Biophys. Acta **1073:** 463–469.

AMY, C. M, B. WILLIAMS-AHLF, J. NAGGERT and S. SOUTH, 1992 Intron-exon organization of the gene for the multifunctional animal fatty acid synthase. Proc. Natl. Acad. Sci. USA **89:** 1105–1108.

AMY, N. K., 1981 Identification of the molybdenum cofactor in chlorate-resistant mutants of *Escherichia coli*. J. Bacteriol. **148:** 274–282.

ARTAVANIS-TSAKONAS, S. M., M. A. T. MUSCAVITCH and B. YEDVOBNICK, 1983 Molecular cloning of *Notch*, a locus affecting neurogenesis in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 1977–1981.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

BAKER, B., 1973 The maternal and zygotic control of development by *cinnamon*, a new mutant in *Drosophila melanogaster*. Dev. Biol. **33:** 429–440.

BENTON, W., and R. DAVIS, 1977 Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. Science **196:** 180–182.

BEVERLEY S. M., and A. C. WILSON, 1984 Molecular evolution in Drosophila and higher Diptera II. A time scale for fly evolution. J. Mol. Evol. **21:** 1–13.

BIESSMANN, H., 1985 Molecular analysis of the *yellow* gene region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **82:** 7369–7373.

BIRNBOIM H., and J. DOLY, 1979 A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:** 1512–1523.

BRAAKSMA, F. J., and W. J. FEENSTRA, 1982 Nitrate reduction in the wild type and a nitrate reductase deficient mutant of *Arabidopsis thaliana*. Physiol. Plant **54:** 351–360.

CAMPUZANO, S., L. CARRAMOLINO, C. V. CABRERA, M. RUIZ-GOMEZ, R. VILLARES *et al.*, 1985 Molecular genetics of the *achaete-scute* gene complex of *D. melanogaster*. Cell **40:** 327–338.

COHEN, H. J., R. T. DREW, J. K. JOHNSON and K. V. RAJAGOPALAN, 1973 Molecular basis of the biological function of molybdenum: the relationship between sulfite oxidase and the acute toxicity of bisulfite and SO_2. Proc. Natl. Acad. Sci. USA **70:** 365–369.

DAVIDSON, J. N., K. C. CHEN, R. S. JANUSON, L. A. MUSMANNO and C. B. KERN, 1993 The evolutionary history of the first three enzymes in pyrimidine biosynthesis. BioEssays **15:** 157–164.

FEINBERG, A., and B. VOGELSTEIN, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

FINNERTY, V., 1976 Genetic units of Drosophila: simple cistrons, pp. 721–765 in *Genetics and Biology of Drosophila*, Vol. 1B, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.

FLEMING, R. J., S. M. DESIMONE and K. WHITE, 1989 Molecular isolation and analysis of the *erect wing* locus in *D. melanogaster*. Mol. Cell. Biol. **9:** 719–725.

FROEHNER, S. C., 1993 Anchoring glycine receptors. Nature **366:** 719.

HENIKOFF, S., M. A. KEENE, J. S. SLOAN, J. BLESKAN, R. HARDS *et al.*, 1986 Multiple purine pathway enzyme activities are encoded at a single genetic locus in Drosophila. Proc. Natl. Acad. Sci. USA **83:** 720–724.

HINTON, S. M., and D. DEAN, 1990 Biogenesis of molybdenum cofactors. Crit. Rev. Microbiol. **17:** 169–188.

HULTMARK, D., R. KLEMENZ and W. J. GEHRING, 1986 Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp* 22. Cell **44:** 429–438.

JACKSON, I., 1991 A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res. **19:** 3795–3798.

JOHNSON, J. L., and S. K. WADMAN, 1989 Molybdenum cofactor deficiency, Chap. 56 in *Inherited Basis of Metabolic Disease*, edited by J. B. STANBURY and J. B. WYNGAARDEN. McGraw-Hill, New York.

JOHNSON, J. L., M. M. WUEBBENS, R. MANDELL and V. SHIH, 1989 Molybdenum cofactor biosynthesis in humans: identification of two complementation groups of cofactor-deficient patients and preliminary characterization of a diffusible molybdopterin precursor. J. Clin. Invest. **83:** 897–903.

JOHNSON, J. L., N. R. BASTIAN and K. V. RAJAGOPALAN, 1990 Molybdopterin guanine dinucleotide: a modified form of molybdopterin identified in the molybdenum cofactor of DMSO reductase from *R. spaeroides*. Proc. Natl. Acad. Sci. USA **87:** 3190–3194.

JOHNSON, M. E., and K. V. RAJAGOPALAN, 1987 Involvement of *chl* A, E, M, and N loci in *E. coli* molybdopterin biosynthesis. J Bacteriol. **169:** 117–125.

KELLER, E. C., and E. GLASSMAN, 1965 Phenocopies of the *ma*-l and *ry* mutants of *Drosophila melanogaster*: inhibition *in vivo* of xanthine dehydrogenase by 4-hydroxypyrazolo-(3,4d)-pyrimidine. Nature **208:** 202–203.

KIRSCH, J., I. WOLTERS, A. TRILLER and H. BETZ, 1993a Gephyrin antisense oligonucleotides prevent glycine receptor clustering in the spinal neurons. Nature **366:** 745–748.

KIRSCH, J., M. L. MALOSIO, I. WOLTERS and H. BETZ, 1993b Distribution of gephyrin transcripts in the adult and developing rat brain. Eur. J. Neurosci. **5:** 1109–1117.

KOZAK, M., 1984 Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:** 857–872.

KRAMER, S. P., J. L. JOHNSON, A. A. RIBEIRO, D. S. MILLINGTON and K. V. RAJAGOPALAN, 1987 The structure of the molybdenum cofactor. Characterization of di-(carboxamidomethyl)-molybdopterin from sulfite oxidase and xanthine oxidase. J. Biol. Chem. **262:** 16357–16363.

LEFEVRE, G., 1981 The distribution of randomly x-ray induced sex-linked genetic effects in *Drosophila melanogaster*. Genetics **99:** 461–480.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, New York.

LIVAK, K., 1984 Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. Genetics **107:** 611–634.

MANIATIS, T., R. C. HARDISON, E. LACY, J. LAUER, C. O'CONNELL *et al.*, 1978 The isolation of structural genes from libraries of eukaryotic DNA. Cell **15:** 687–701.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK 1982 *Molecular Cloning*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MEIDINGER, R. G., and M. M. BENTLEY, 1986 Genetic and Developmental Characterization of the *Aldox*-2 locus of *Drosophila melanogaster*. Biochem. Genet. **24:** 683–699.

MOERMAN, D. G., G. M. BENIAN, R. J. BARSTEAD, L. A. SCHRIEFER and R. WATERSTON, 1988 Identification and intracellular localization of the unc-22 gene product of *C. elegans*. Genes Dev. **2:** 93–105.

NASON, A., K. Y. LEE, S. PAN, P. A. KETCHUM, A. LAMBERTI *et al.*, 1971 *In vitro* formation of assimilatory reduced NADP:nitrate reductase from a *Neurospora* mutant and a component of molybdenum-enzymes. Proc. Natl. Acad. Sci. USA **68:** 3242–3246.

NOHNO, T., Y. KASAI and T. SAITO, 1988 Cloning and sequencing of the *E. coli chl*EN operon involved in molybdopterin biosynthesis. J. Bacteriol. **170:** 4097–4102.

PADILLA, H. M., and W. G. NASH, 1977 A further characterization of the *cinnamon* gene in *Drosophila melanogaster*. Mol. Gen. Genet. **155:** 171–177.

PATEMAN, J. A., D. J. COVE, B. M. REVER and D. B. ROBERTS, 1964 A common cofactor for nitrate reductase and xanthine dehydrogenase which also regulates the synthesis of nitrate reductase. Nature **201:** 58–60.

PERRIMON, N., L. ENGSTROM and A. P. MAHOWALD, 1989 Zygotic lethals with specific maternal effect phenotypes in *Drosophila melanogaster*. I. Loci on the *X* chromosome. Genetics **121:** 333–352.

PITTERLE, D. M., J. L. JOHNSON and K. V. RAJAGOPALAN, 1993 *In vitro* synthesis of molybdopterin from precursor Z using purified converting factor. Role of protein-bound sulfur in formation of the dithiolene. J. Biol. Chem. **268:** 13506–13509.

PRIOR, P., B. SCHMITT, G. GRENNINGLOH, I. PRIBILLA, G. MULTHAUP *et al.*, 1992 Primary structure and alternative splice variants of gephyrin, a putative glycine receptor-tubulin linker protein. Neuron **8:** 1161–1170.

RAJAGOPALAN, K. V., and J. L. JOHNSON, 1992 The pterin molybdenum cofactor, a minireview. J. Biol. Chem. **267:** 10199–10202.

RIVERS, S. L., E. MCNAIRN, F. BLASCO, G. GIORDANO and D. H. BOXER, 1993 Molecular genetic analysis of the *moa* operon of *Escherichia coli* K12 required for molybdenum cofactor biosynthesis. Mol. Microbiol. **8:** 1071–1081.

SANGER, F., S. NICKLEN and A. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHALET, A. P., 1986 The distribution of and complementation relationships between spontaneous x-linked recessive lethal mutations recovered from crossing long-term laboratory stocks of *Drosophila melanogaster*. Mutat. Res. **163:** 115–144.

SCHOTT, D. R., M. C. OLSON and V. FINNERTY, 1986 Molybdenum hydroxylases in *Drosophila*. III. Characterization of the low xanthine dehydrogenase gene. Biochem. Genet. **24:** 509–527.

SCHWARTZ, R. M., and M. O. DAYHOFF, 1979 Dayhoff table, pp. 353–358 in *Atlas of Protein Sequence and Structure*, edited by M. O. DAYHOFF. National Biomedical Research Foundation, Washington D.C.

SHANMUGAM, K. T., V. STEWART, R. P. GUNSALUS, D. H. BOXER, J. A. COLE *et al.*, 1992 Proposed nomenclature for the genes involved in molybdenum metabolism in *Escherichia coli* and *Salmonella typhimurium*. Mol. Microbiol. **6:** 3451–3454.

SHELTON, M. E., 1990 Molecular analysis of the *cinnamon* locus in *Drosophila melanogaster*. M.S. Thesis, Emory University, Atlanta, Ga.

SIMMONDS, H. A., S. REITER and T. NISHINO, 1994 Hereditary Xanthinuria, Chap. 54 in *Metabolic Basis of Inherited Disease*, Ed. 7. McGraw-Hill, New York (in press).

SINDHU, R. K., D. H GRIFFIN and D. C. WALTON, 1990 Abscisic aldehyde is an intermediate in the enzymatic conversion of xanothoxin to abscisic acid in *Phaseolus vulgaris* L. leaves. Plant Physiol. **93:** 689–694.

STEWART, V., and C. H. MACGREGOR, 1982 Nitrate reductase in *E. coli* K-12: involvement of *chl*E, *chl*F and *chl*G loci. J. Bacteriol. **151:** 788–799.

STIVALETTA, L. A., C. K. WARNER, S. LANGLEY and V. FINNERTY, 1988 Molybdoenzymes in *Drosophila*. IV. Further characterization of the *cinnamon* phenotype. Mol. Gen. Genet. **213:** 505–512.

TASAYCO, M. J., and G. D. PRESTWICH, 1990 Aldehyde oxidases and dehydrogenases in antennae of five moth species. Insect Biochem. **20:** 691–700.

WAHL, R. C., C. K. WARNER, V. FINNERTY and K. V. RAJAGOPALAN, 1982 *Drosophila melanogaster ma-l* mutants are defective in the sulfuration of desulfo Mo hydroxylases. J. Biol. Chem. **257:** 3958–3962.

WARNER, C. K., and V. FINNERTY, 1981 Molybdenum hydroxylases in *Drosophila*. II. Molybdenum cofactor in xanthine dehydrogenase, aldehyde oxidase and pyridoxal oxidase. Mol. Gen. Genet. **184:** 92–96.

Warner, R. L., and A. KLEINHOFS, 1992 Genetics and molecular biology of nitrate metabolism in higher plants. Physiol. Plant. **85:** 245–252.

WUEBBENS, M. M., and K. V. RAJAGOPALAN, 1993a A Study of the Initial steps of Molybdopterin Biosynthesis in *Escherichia coli*, p. 163 in *Proceedings of the Tenth International Symposium on Chemistry and Biology of Pteridines and Folates*, edited by J. AYLING, G. NAIR and C. BAUGH. University of South Alabama College of Medicine, Mobile.

WUEBBENS, M. M., and K. V. RAJAGOPALAN, 1993b Structural characterization of a molybdopterin precursor. J. Biol. Chem. **268:** 13493–13498.

YAMAMOTO, K., B. ALBERTS, R. BENZINGER, L. LAWHOME and G. TREIBER, 1970 Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large scale virus purification. Virology **40:** 734–744.

Communicating editor: R. E. DENELL